

Two thresholds are better than one

Tao Zhang

Terrance E. Boulton

R.C. Johnson

Vision And Security Technology Lab
University of Colorado
Colorado Springs, CO 80918

{zhangt, tboulton, rcjohnson}@vast.uccs.edu

Abstract

The concept of the Bayesian optimal single threshold is a well established and widely used classification technique. In this paper, we prove that when spatial cohesion is assumed for targets, a better classification result than the “optimal” single threshold classification can be achieved. Under the assumption of spatial cohesion and certain prior knowledge about the target and background, the method can be further simplified as dual threshold classification. In core-dual threshold classification, spatial cohesion within the target core allows “continuation” linking values to fall between the two thresholds to the target core; classical Bayesian classification is employed beyond the dual thresholds. The core-dual threshold algorithm can be built into a Markov Random Field model (MRF). From this MRF model, the dual thresholds can be obtained and optimal classification can be achieved. In some practical applications, a simple method called symmetric subtraction may be employed to determine effective dual thresholds in real time. Given dual thresholds, the Quasi-Connected Component algorithm is shown to be a deterministic implementation of the MRF core-dual threshold model combining the dual thresholds, extended neighborhoods and efficient connected component computation.

1. Introduction

The segmentation or detection of a “target” of interest within an image is a common problem spanning the areas of video surveillance, medical imaging, machine vision, document processing, microscopy, and many other applications. There have been hundreds of papers in the area of thresholding, e.g. see the reviews in [15, 7]. While the papers vary in how to compute the threshold, all these papers have one thing in common – they all define a single threshold. A single threshold provides a fast and convenient segmentation and is generally easy to analyze. There

have been multiple “optimal” algorithms differing in their assumptions and criterion. The most commonly used of these is the optimal Bayesian formulation, which will be formally discussed in the next section. Other criterion have been used for “optimal” thresholds based on histogram shape, clustering-based techniques, entropy-based techniques, attributes-based techniques, and locally adaptive thresholding.

The most relevant past work to this paper are those that mixed spatial information into the process of determining the threshold. One of the earliest of these was [10], which used local window averages. Others have used windows of local variances and second order statistics[1], edge features[17], quad-trees[18] and hysteresis threshold selection[14]. If we had enough information, the ideal would be to apply the Bayesian thresholding to full joint spatial/intensity distribution models. This joint distribution can be used to define a per-pixel probability, that mixes space and intensity data, and one could apply a single threshold on that probability to segment the data. But knowledge of a full joint distribution is an unrealistic assumption. Weaker models have been developed as approximation, e.g. using co-occurrence [12]. These approximate techniques are still quite expensive (generally quadratic) and in experimental evaluations, e.g. [15], have been found to be significantly less effective.

However, not all thresholding papers use a single threshold. An early dual-thresholding approach was described in [9], wherein models of background using “known background pixels” are used to define an upper and lower threshold for the backgrounds distributions such that targets with values near the average background value can still be segmented. This approach, and many since then, use multiple thresholds to address multi-modal backgrounds, wherein the target is within the area. This is well known and, of course, useful. But with a bit of training data this is really just separating the problem of threshold selection to above and below the target region and can be applied to extend the approach presented herein, moving from a core-dual thresh-

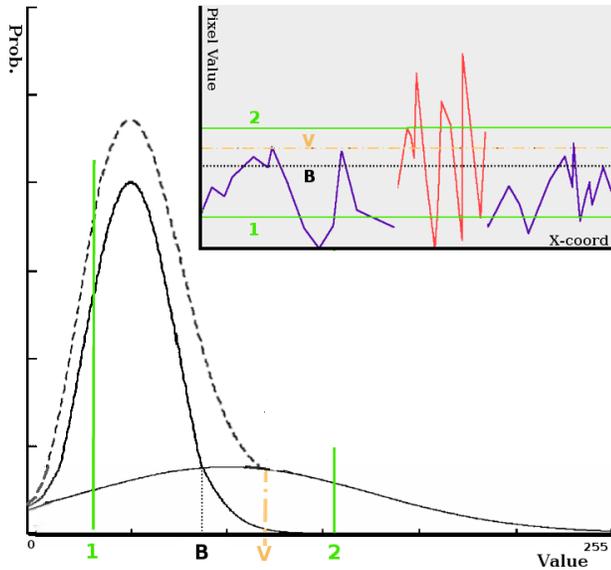


Figure 1. Comparison of single and core-dual thresholds. The solid broad curve is the probability distribution of target values and the solid taller curve is the probability distribution of background values. The dashed curve is the observed total distribution for 1/4 of all pixels being on target. The dotted black line, marked B, is the Bayes optimal single threshold. The orange (very light grey) dot-dashed line, marked V, is the optimal “valley” single threshold. The green (medium gray) solid lines, marked 1 and 2, are examples for a core-dual threshold. The gray rectangular insert is an hypothetical scan-line showing the target in red (dark grey) and background in blue (almost black) as well as the 4 thresholds.

old approach to a core-quad threshold approach.

The idea of dual thresholds, with hysteresis, goes back to the Canny edge detector [5]. That work, however, was just for edge detection, using the edge strength thresholds along the prescribed curve boundary, and provided no formal method for defining the thresholds. Latter Hancock and Kittler [8] derived, under the assumption of Gaussian noise, a formal approach for estimating the thresholds in a Canny edge detector. Besides the strong model assumption, their work is strongly tied to the Canny edge detector. While a similar issue, the overall approach does not apply to threshold selection for 2D region detection.

A different approach to dual thresholding for region-based detection is presented in [4], and is introduced as a way of balancing spatial information with standard thresholding in a real-time system. The basic idea of the dual threshold can be explained in Figure 1. In the figure, all target pixels above threshold 1, which are connected to some pixel above threshold 2, would be labeled target. This correctly labels all of the background, and almost all of the target with only very small parts of the target missing. Given the probability of targets and background values, Bayes op-

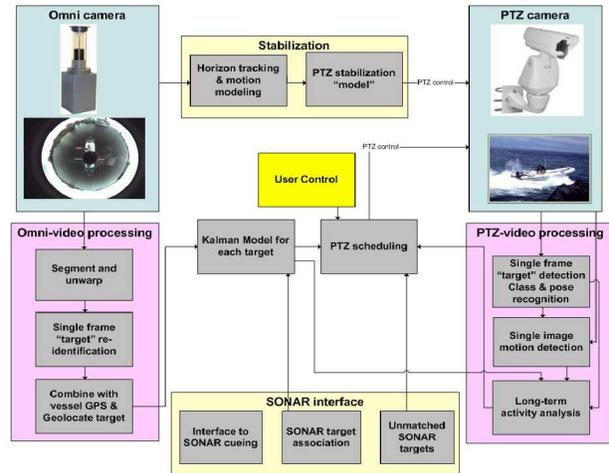


Figure 2. Application behind the dual-threshold selection problem, detection of small targets in the omni-directional image followed by detection and tracking in the PTZ.

timal chooses a single threshold where the two probabilities are equal and, in the example, produces significant missed pixels in the target as well as false alarms in the background. The valley algorithm produces fewer false alarms but also fails to detect most of the target. The intuition of this approach, exploiting the spatial cohesion of targets, allows a much lower second threshold that fills in much of the target region. But the requirement of connecting to the higher threshold stops spurious background regions from forming. However, [4] developed/presented the technique in an ad hoc manner and argued heuristically that it is better. They never provided any formal analysis and determined thresholds heuristically from ROC curves of system performance. That approach may apply for a static camera, doing background subtraction, in a static scene. But region detection and dual-thresholding can be applied in a wider range of video surveillance and vision problems, not just background subtraction. E.g. if looking for “hot” targets in thermal video, they can be applied directly to the intensity imagery. With statistical models or texture models, one can produce a per-pixel target filter response image (ideally a likelihood), and then apply the dual-thresholding approach to the filter response.

Our underlying problem is detection and tracking of small vessels from a moving ship, with an omni-directional camera as the initial detection platform and a PTZ as an assessment image. The overall flow, which is not covered in this paper, is show in figure 2.

Rather than focusing on the application, the focus of this paper is to take the dual-thresholding with spatial cohesion idea, formalize it and develop techniques for estimating thresholds from the images themselves. In particular we

formally prove, that for spatially cohesive targets, it is more effective than single Bayesian threshold. The proof technique can be applied for just about any definition of a single “optimal” threshold. We then develop approaches for determining the thresholds. If the background/foreground models are known, or can be estimated, then setting the thresholds is formulaic. We present an MRF approach when one can determine thresholds a priori, and a symmetric subtraction based technique that can be applied per image when a little a priori data can be used. Either technique may be used to determine thresholds every few hundred frames, and hence adapt the thresholds over time.

2. Problem Formation

2.1. Bayesian Optimal Classification

We first consider the two class pattern classification problem. Let ω be the category state, $\omega = \omega_1$ be the target, $\omega = \omega_2$ be the background. Let $x \in [0, T_{max}]$ be a continuous feature variable. Using Bayesian theorem, we can compute the posterior probability from the prior probability $P(\omega)$ and the likelihood density $p(x|\omega)$ as:

$$P(\omega|x) = \frac{p(x|\omega)P(\omega)}{\sum_{\omega} p(x|\omega)P(\omega)} \quad (1)$$

Let the overall risk for taking x as a decision threshold be $\alpha(x)$. The Bayesian optimal threshold is $x_{opt} = \arg \min_x \alpha(x)$.

If we have more knowledge on the problem, better results may be achieved. Without loss of generality, we assume that there is only one x_{opt} and the mean of feature values for the target are smaller than that of the background. The following discussion can be easily generalized to cases where multiple x_{opt} exist and where the mean of feature values of the background are smaller.

When zero-one loss function is employed as the risk function, the Bayes risk [6] is

$$\alpha(x_{opt}) = \int_0^{x_{opt}} P(\omega_2|x)dx + \int_{x_{opt}}^{T_{max}} P(\omega_1|x)dx \quad (2)$$

2.2. Optimal Classification with Spatial Knowledge

Besides the feature value likelihood distribution, if we further know the spatial likelihood distribution of the target and the background, the joint likelihood probability $P(\omega|i, x)$ is known, where i is the position index, then bet-

ter results α_2 can be achieved in terms of risk:

$$\begin{aligned} \alpha_2 &= \int_0^{x_{opt}} \sum_i \min(P(\omega_2|i, x), P(\omega_1|i, x))dx \\ &+ \int_{x_{opt}}^{T_{max}} \sum_i \min(P(\omega_2|i, x), P(\omega_1|i, x))dx \\ &\leq \int_0^{x_{opt}} \sum_i P(\omega_2|i, x)dx + \int_{x_{opt}}^{T_{max}} \sum_i P(\omega_1|i, x)dx \\ &= \alpha(x_{opt}) \end{aligned} \quad (3)$$

In the computation of risk $\alpha(x_{opt})$, for a given x , the decisions are the same regardless of the location of x ; with the introduction of spatial likelihood probability, the decision is made at each position i and results in better classification.

It may happen that for some fixed T_1 , $P(\omega_1|i, x) > P(\omega_2|i, x)$ holds for most $x < T_1$ regardless of i . In this case, a classical decision can be made for $x < T_1$ without considering the position to get a better result. This generally reflects our knowledge of the target. For example, we may know that if $x < T_1$, the points belong to the target. We may have similar knowledge on the background. For some T_2 , if $x > T_2$, the points belong to the background.

If such knowledge about T_1 and T_2 is available, we can keep the classical decision rule for $x < T_1$ and T_2 and only consider spatial information for $x \in [T_1, T_2]$. The risk α_3 by doing so is

$$\begin{aligned} \alpha_3 &= \int_0^{T_1} P(\omega_2|x)dx + \int_{T_2}^{T_{max}} P(\omega_1|x)dx \\ &+ \int_{T_1}^{T_2} \sum_i \min(P(\omega_2|i, x), P(\omega_1|i, x))dx \end{aligned} \quad (4)$$

Obviously, $\alpha_2 < \alpha_3 < \alpha(x_{opt})$. The decision process is simplified and a good result is achieved. However, the determination of dual thresholds is an ill-posed problem. We will show how to determine the core-dual thresholds.

3. An Example

We consider the following case as an example. The prior probabilities for target and background are the same $P(\omega_1) = P(\omega_2) = 0.5$. The likelihood probability distributions for target $P(x|\omega_1)$ and background $P(x|\omega_2)$ are $N \sim (\mu_1, \sigma_1^2)$ and $N \sim (\mu_2, \sigma_2^2)$ respectively. We also assume the spatial likelihood distribution $P(\omega_1|i)$ of the target is known, as shown in equation (5). This implies the spatial likelihood distribution of background $P(\omega_2|i) = 1 - P(\omega_1|i)$ since only two classes are involved. The feature variable is x and the position index is i .

$$P(\omega_1|i) = \frac{1}{1 + \frac{\|pos(i) - \mu_p\|^2}{(9\sigma_p)^2}} \quad (5)$$

where μ_p is the center of the target, σ_p is the radius of the target, $pos(i)$ is the coordinate at position i .

With the spatial information, the decision rule is to classify the pixel as ω_1 if $P(\omega_1|x, i) > P(\omega_2|x, i)$, otherwise ω_2 at each position. The discriminant function is:

$$\begin{aligned} G(x, i) &= P(\omega_1|i, x) - P(\omega_2|i, x) \\ &\propto P(x|\omega_2) - P(\omega_1|i)(P(x|\omega_1) + P(x|\omega_2)) \end{aligned} \quad (6)$$

If $P(\omega_1|i) = 0.5$, positions provide no information and the discriminant function is reduced to the classical function. Otherwise, this discriminant function is position dependent. For the test image shown in Figure 3, we set $\mu_1 = 90$, $\mu_2 = 150$, $\sigma_1 = \sigma_2 = 30$, $\mu_p = (250, 250)$, $\sigma_p = 40$. The optimal single Bayesian threshold is $x_{opt} = \frac{\mu_1 + \mu_2}{2} = 120$.

3.1. Ratio Map

Since the likelihood probabilities and the spatial likelihood probability are known, a classification decision can be made at every point i . This is the optimal classification given the spatial likelihood distribution. From the optimal classification, we can define the ratio map T_{map} .

$$T_{map}(x) = \frac{\sum_i H(G(x, i))}{\sum_i \delta(x_i - x)} \quad (7)$$

where x_i is the feature value at i , $\delta(\cdot)$ is Delta function and $H(\cdot)$ is Heaviside function,

$$H(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (8)$$

Ratio map $T_{map}(x)$ is the ratio of the number of positions whose feature values are x that are classified as targets to the total number of positions whose feature values are x . Figure 3 shows the ratio map for the optimal classification of the test image.

The ratio map clearly shows two points. For the left point, all pixels whose feature value is less than that point will be exclusively classified as target points; for the right point, all pixels whose feature value is greater than that point will be exclusively classified as background points. The selection of these two points is not unique. These two points reflect our knowledge about the likelihood probability of feature variables for the target and the background when feature values are beyond these two points. The ratio map provides a way to determine core-dual thresholds.

3.2. Optimal Core-Dual Thresholds

Some restrictions may be applied so that the selection of core-dual thresholds given a ratio map can be unique. We can define T_1 and T_2 as:

$$T_1 = \arg \max_t T_{map}(t) > \zeta_1 \quad for \quad x \leq t \quad (9)$$

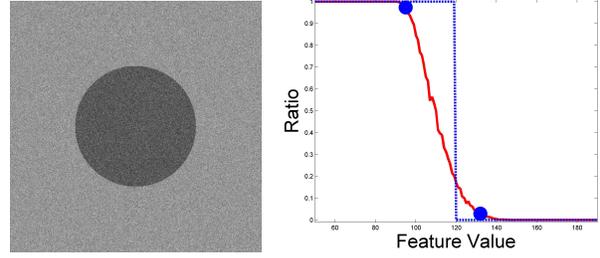


Figure 3. Left is a simple noisy test image. Right is the ratio maps from optimal dual threshold (solid line) and optimal single threshold classification (dotted line). $x_{opt} = 120$, $T_1 = 95$ and $T_2 = 132$. The small dots indicate positions of dual thresholds when $\zeta_1 = 0.96$ and $\zeta_2 = 0.04$.

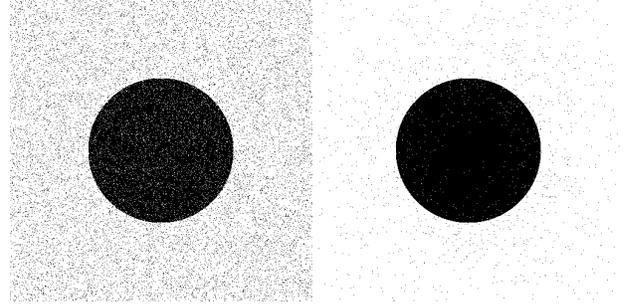


Figure 4. Classification results by single threshold method and core-dual threshold method. Left is Bayesian Optimal Solution. The right is core-dual with $\zeta_1 = 0.96$, $\zeta_2 = 0.04$. The classification error rates are 6.7%, 0.76% respectively. Note it NOT doing full spatial filtering using knowledge of expected target size, adding those techniques (later sections of rapper) produces near perfect segmentation.

$$T_2 = \arg \min_t T_{map}(t) < \zeta_2 \quad for \quad x > t \quad (10)$$

where ζ_1 is a number close to 1 and ζ_2 is a number close to 0. They reflect our confidence in the choice of T_1 and T_2 . If $\zeta_1 = 0.96$, $\zeta_2 = 0.04$, we can determine the optimal core-dual thresholds as $T_1 = 95$, $T_2 = 132$ from the ratio map.

We showed in section 2.2 that if spatial likelihood distribution is known, then dual threshold method will achieve better classification results. This conclusion is confirmed experimentally, as shown in Figure 4.

Figure 5 shows the ROC curves for the single threshold case, as well as for varying each of the dual thresholds. The core-dual thresholds are clearly superior in the example. The classification errors do not change for a large range of feature values. This fact tells us the selection of core-dual thresholds is not unique but also that the results is not terribly sensitive to minor perturbations once in the right region. One goal of the definition of optimal dual thresholds is to make the selection unique.

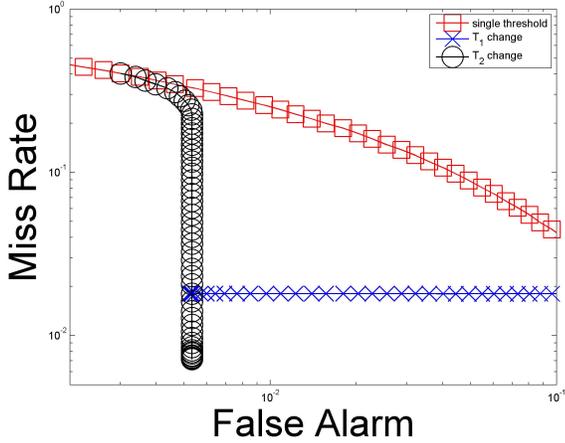


Figure 5. ROC curves for single threshold and for each dual threshold while holding the other constant at its optimal setting for $\zeta_1 = 0.96$, $\zeta_2 = 0.04$.

4. Markov Random Field Model

In previous sections, the spatial likelihood probability for targets is assumed to be known. In real problems, such assumption is seldom true. Instead of the assumption of known spatial likelihood probability, we choose to model the spatial cohesion in the dual threshold method. Spatial cohesion is a reasonable assumption in practical problems. Markov random field (MRF) provides a principled way to model local spatial distributions.

The core-dual threshold method can be modeled as a MRF because the classification of a pixel depends on its neighbor pixels. Let site S represent the image lattice with size $m \times n$, the label set be $L = \{\omega_1 = 1, \omega_2 = -1\}$, $f = \{f_1, f_2, \dots, f_{m \times n}\}$ be a family of random variables defined on S and takes a value in L . From the equivalence of MRF and Gibbs Random Field [2], the distribution takes the following form.

$$P(f) = Z^{-1} \times e^{-U(f)} \quad (11)$$

$Z = \sum_f e^{-U(f)}$ is a normalizing constant, $U(f) = \sum_C V_C(f)$ is the energy function and $V_C(f)$ is the potential function over clique C . In the maximum a posteriori (MAP) framework, given observation d , the posterior energy function $U(f|d)$ is the summation of prior energy $U(f)$ and the likelihood energy $U(d|f)$ [11].

$$U(f|d) = U(f) + U(d|f) \quad (12)$$

4.1. The Prior Model

The spatial cohesion can be modeled by the Ising model[11]. The prior energy is given by:

$$U(f) = \sum_{(i,i') \in N} \beta_1 f_i f_{i'} \quad (13)$$

where N is the 4 nearest neighbors, $\beta_1 > 0$ controls the weight of the prior energy. The Ising model gives the boundary length of targets. If the targets are cluttered, the boundary length and thus the prior energy tend to be large; if the targets are in blob shape, the boundary length and its corresponding energy tend to be small. (Note alternative 8 or even 16 connected neighbors[19] could be considered but would be slower.) The optimization of this prior energy encourages spatial cohesion.

4.2. The Observation Model

The likelihood probability given an observation d at each site S depends not only on the labeling of neighbor sites, but also on the dual thresholds. It should reflect both the spatial cohesion assumption and our knowledge on the feature distributions for the target and the background beyond the dual thresholds. A simple model such as the Gaussian distribution cannot achieve the goal. The Gaussian assumption may not be true in many real situations and it can not incorporate our knowledge on the dual thresholds.

We only consider the likelihood involving at most pair site cliques, and will define the single site clique energy and pair site energy respectively.

- Single Site Energy

The single site potential energy is defined as:

$$V_{C_1} = \sum_{i \in S} A(d) \cdot \beta_2 \cdot f_i \quad (14)$$

$$A(d) = H(d - T_1) + H(d - T_2) - 1 \quad (15)$$

where C_1 are single site cliques, $\beta_2 > 0$ control the weight of single site energy. $A(d)$ is a feature value selection function. $A(d)$ is nonzero only on points with feature values $d \in [0, T_1] \cup [T_2, T]$. Under this definition, if $d \in [0, T_1]$, minimization of the energy encourages $f_i = \omega_1$; if $d \in [T_1, T]$, $f_i = \omega_2$ is encouraged. The single site energy perfectly incorporates our knowledge of the distribution of the target and background beyond T_1 and T_2 .

- Pair Site Energy

The pair site energy is defined as:

$$V_{C_2} = \sum_{(i,i') \in C_2} B(d) \cdot f_i \cdot f_{i'} \cdot g(i') \quad (16)$$

$$B(d) = H(d - T_2) - H(d - T_1) \quad (17)$$

$$g(i') = \begin{cases} \zeta & f_{i'} = 1 \\ \eta & f_{i'} = -1 \end{cases} \quad (18)$$

where C_2 are pair site cliques. $B(d)$ is nonzero only on points with feature values $d \in [T_1, T_2]$. $g(\cdot)$ is a function that defines the interaction between site i and i' . The pair site energy encourages spacial cohesion. $\zeta, \eta \geq 0$ controls the preference for targets or background. $\zeta > \eta$ favors the target; $\zeta < \eta$ favors the background; $\zeta = \eta$ favors the majority in the neighbor region.

The observation model is

$$U(d|f) = \sum_{i \in S} A(d) \cdot \beta_2 \cdot f_i + \sum_{(i,i') \in C_2} B(d) \cdot f_i \cdot f_{i'} \cdot g(i') \quad (19)$$

The observation model may be treated as a special auto-model[11] with selection functions based on feature values.

4.3. Parameter Estimation

With known dual thresholds, the energy can be minimized by ICM[3] or simulated annealing[16]. In this paper, we use simulated annealing with a Metropolis sampler[13]. The more interesting problem is how to estimate the dual thresholds.

The optimal core-dual threshold determined by formulas (9) and (10). The final ratio map will be similar to Figure 3. In that case, the threshold T_1 (T_2) can be computed by gradient ascent (descent) as in equation (20) and (21). Otherwise, more measures are needed to update the thresholds in the correct direction.

We will explain how to update T_1 in these cases. From the design of the energy functions, it is obvious that the values in the ratio map for $t < T_1$ are expected to be large when the energy is minimized. If this is the case, and the ratio map curve around T_1 is flat, the value of T_1 is an acceptable one. However, we update T_1 in increasing direction to ensure the uniqueness of T_1 ; if the values are very small instead, T_1 is selected too small. We call this T_1 unreliable. Similar analysis applies to updating T_2 : when the values in the ratio map for $t > T_2$ are too large, the selection of T_2 is too large. We call this T_2 unreliable. T_1 and T_2 may not be unreliable at the same time.

$$T_1^{i+1} = T_1^i + \mu_1 \nabla T_{map}(T_1^i) \quad (20)$$

$$T_2^{i+1} = T_2^i - \mu_2 \nabla T_{map}(T_2^i) \quad (21)$$

where $\mu_1, \mu_2 > 0$ are step sizes.

The quality of classification is measured by the prior energy (13). Smaller energy generally corresponds to better classification. Unfortunately, it is not always true. For example, when $T_1 = 10$ and $T_2 = 25$, the prior energy is 0

because no pixels are classified as targets; when $T_1 = 82$ and $T_2 = 137$, the prior energy is 3968. The classification result is acceptable only if the prior energy is low and both T_1 and T_2 are reliable. When one of T_1 and T_2 is unreliable, the prior energy is unreliable and the classification result is unacceptable.

The dual threshold estimation algorithm in one iteration is summarized below:

ParameterEstimation(T_1, T_2)

1 $T_{reliable1} = \text{true};$
 $T_{reliable2} = \text{true};$

2 Estimation

2.1 $T_s = \frac{T_1 + T_2}{2}$. Initial classification using T_s .

2.2 Energy minimization of (12) using Simulated annealing given T_1 and T_2 . The result is $Result_m$

2.3 Compute ratio map T_{map} and prior energy E_{prior} .

(a) IF $T_{map}(T_1) < \xi_1, T_{reliable1} = \text{false}$
 IF $T_2 - T_1 < 2, T_2 = T_2 + 1$
 ELSE $T_1 = T_1 + 1$

(b) IF $T_{map}(T_2) > \xi_2, T_{reliable2} = \text{false}$
 IF $T_2 - T_1 < 2, T_1 = T_1 - 1$
 ELSE $T_2 = T_2 - 1$

(c) Update T_1 and T_2 by equation (20) and (21).

ξ_1, ξ_2 reflect our confidence on the knowledge about the distribution of targets and backgrounds beyond T_1 and T_2 . For example, we can set $\xi_1 = 0.98, \xi_2 = 1 - \xi_1$.

The simultaneous minimization of the energy and the estimation of the dual thresholds is given below:

EnergyMinimization

1. Initialization $T_1^0, T_2^0, E_{min}, Result_m$ and $Trust = \text{false}$

2. DO $i \leftarrow i + 1$

- $Trust = \text{false}$
- $(T_1^{i+1}, T_2^{i+1}) \leftarrow \text{ParameterEstimation}(T_1^i, T_2^i)$
- IF T_1^{i+1} and T_2^{i+1} are both reliable,
 update $Result_m,$
 $Trust = \text{true},$
 $E_{min} = \min(E_{prior}, E_{min})$

UNTIL $i > i_{max}$ OR $(T_1^{i+1} = T_1^i, T_2^{i+1} = T_2^i)$

3. Check validity

- IF $Trust = \text{false}$ OR IF $E_{prior} \gg E_{min}$
 No object detected.
- ELSE accept the result

The algorithm is robust to initialization. The dual thresholds are only affected by the confidence ξ_1 and ξ_2 .

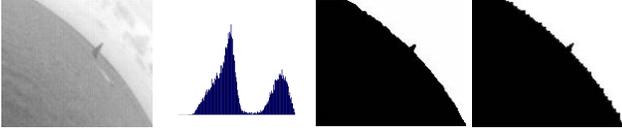


Figure 6. An image captured by a surveillance camera at sea and segmentation results. Left to right: original image containing two small ships, its histogram (sky contribution is significant), segmentation results by the MRF model and QCC when $T_l = 195$, $T_h = 212$. Note some regions of sky are darker than some of the water and a single threshold cannot segment the scene.

5. Deterministic Energy Minimization

Above algorithms provide a principled way to estimate the dual thresholds and minimize the posterior energy simultaneously. The stochastic energy minimization by simulated annealing is slow. In many applications, the scene content is slowly changing and we can justify the use of thresholds over a period of time. But applying the thresholds in applications like surveillance, target or people tracking, must be real time.

In [4], a run time algorithm called Quasi-Connected Components (QCC) was proposed as a grouping method for a visual surveillance and tracking system. QCC assumes the low threshold T_l and the high threshold T_h . An image pixel whose feature value is greater than the high threshold is classified as a target pixel. For image pixels with feature values between T_l and T_h , they are classified as target pixels only if they are quasi-connected with a region containing at least one pixel whose feature value is greater than the high threshold. The quasi-connected components differ from traditional connected components in that the computation is carried out using a generalized neighborhood similar to N_{16} [19], but more efficient because it uses a reduced image. This multi-scale scheme improves the computation speed and the generalized neighborhood automatically fills small gaps. From the description of QCC, it appears to achieve the minimization of the posterior energy (12). It is a deterministic energy minimization method given the dual thresholds with the gap filling ability.

Unfortunately, Boulton et. al. did not give a theoretical analysis for the method, nor propose a systematic way to determine the dual thresholds. The MRF model proposed in this paper provides a theoretical foundation for the method. We also give an algorithm based on ratio maps to determine the core-dual thresholds. In the MRF model, we need to set $\zeta < \eta$ in equation (18) because the QCC algorithm naturally favors the background.

Figure 6 shows an image captured by a shipboard camera, and segmentation results by the MRF model and QCC given the same dual thresholds. The segmentation results using these two methods are very similar. However, QCC

is much faster, supporting real-time segmentation and tracking. Operationally one can use the MRF running at a slower frame rate to dynamically determine the core-dual thresholds for QCC, and use the QCC on a per-frame basis. We are currently looking at using a Kalman filter to “track” the MRF-produced thresholds and provide for smoother transitions between estimations. A second issue for this is temporally adapting the “window” used in threshold estimation, especially in any image where apparent target sizes vary rapidly over the image, as is the case in our omni-directional image.

6. Symmetric Subtraction

In this section, we give a histogram based method for estimating dual thresholds directly. The histogram can be represented as a mixture model:

$$P(x) = P(\omega_1)p(x|\omega_1) + P(\omega_2)p(x|\omega_2) \quad (22)$$

where $P(\omega_2) = 1 - P(\omega_1)$ and

$$P(\omega_1) = \frac{Size(target)}{Size(region)} \quad (23)$$

The proposed MRF model assumes the trailing distribution beyond T_{max} is ignorable. Otherwise, they will be classified as the background. In Figure 6, our targets are small ships, but the sea is classified as the target due to the significant trailing distribution from the sky. The solution is to sample in small regions, such as 4-8 times of the target size. For many problems, the estimated size of targets is known. From equation 23, smaller regions result in more noticeable contribution from small targets. Also, the histogram from the background tends to be an approximately symmetric single mode (see histograms in Figure 6 and Figure 7) in small regions.

When the histogram from the background can be treated as symmetric, and half of the histogram is known only from background, then we can subtract the background histogram from the whole histogram as defined by equation (22), the remaining histogram is only from the target. The symmetric subtraction algorithm is given below:

SymmetricSubtraction

- 1 Compute Histogram for window as $P(x)$
- 2 Detect the dominant peak as background mean μ_2
- 3 Get the right background histogram $P_2^r(x)$
- 4 Fold $P_2^r(x)$ as the left background histogram $P_2^l(x)$
- 5 Subtract P_2 from $P(x)$ as target histogram $P_1(x)$
- 6 Compute the mean of target as μ_1
- 7 Compute variances for background σ_2 and target σ_1

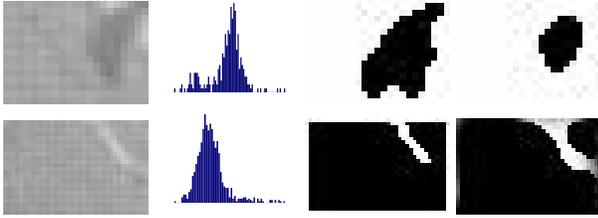


Figure 7. Detection of ships on sea. Left column is enlarged ship images, second are their histograms, third column are their segmentation result by the MRF model, right column are the segmentation results by symmetric subtraction and QCC.



Figure 8. Detection of cars in thermal image. Left to right: original image, detection results from the MRF model (starts from $T_1 = 100$, $T_2 = 200$ and stops at $T_1 = 138$, $T_2 = 152$). Right is detection results with QCC.

$$8 \quad T_1 = \mu_1 - \lambda\sigma_1, T_2 = \mu_2 + \lambda\sigma_2, \lambda > 0$$

The relationship between dual thresholds and classification errors depends on the concrete forms of spatial likelihood distribution, which is unknown for real problems. So the steps in determining the core-dual thresholds are often heuristic. The underlying assumption is that pixels with feature values less than the mean feature value of the target are more likely to be target pixels. A similar assumption is applied to the background. Figure 7 and Figure 8 show the detection of very small targets using the MRF model and symmetric subtraction. Symmetric subtraction works well when the sampling region is small and will often fail if the background is significantly multi-modal within the window.

7. Conclusion

This paper showed that two thresholds are better than one. Because the two thresholds are applied only in the direct neighborhood of the target, i.e. near its core, we call this approach core-dual thresholding. The core-dual thresholding allows effective use of the weak constraint of spatial cohesion, and with only mild assumptions, is provably better than just applying the single “optimal” Bayesian threshold. The paper shows how to use the MRFs to simultaneously segment and determine the optimal values for the core-dual thresholds. It then explored QCC approximations that are much faster than MRFs, and which are suitable to real time implementation and discussed how to use them to set the MRF thresholds for a real-time QCC implementation. Two real examples demonstrated the approaches.

8. Acknowledgments

The work is supported in part by ONR Grant N00014-05-1-0543 and ONR SBIR N00014-05-0020.

References

- [1] N. Ahuja and A. Rosenfeld. A note on the use of second-order gray-level statistics for threshold selection. *IEEE Trans. Syst. Man Cybern.*, SMC-5:383–388, 1975.
- [2] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B*, 36:192–236, 1974.
- [3] J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society, Series B*, 48:259–302, 1986.
- [4] T. Boulton, R. Micheals, X.Gao, and M. Eckmann. Into the woods: Visual surveillance of noncooperative and camouflaged targets in complex outdoor settings. *Proceedings of the IEEE*, 89(10):1382–1402, 2001.
- [5] J. Canny. A variational approach to edge detection. In *Proceedings of AAAI*, 1983.
- [6] R. O. Duda, P. E. Hart, and D. G. Stork. Pattern classification, second edition, 2001. John Wiley & Sons, Inc.
- [7] C. A. Glasbey. An analysis of histogram-based thresholding algorithms. *Graph. Models Image Process.*, 55, 1993.
- [8] E. R. Hancock and J. Kittler. Adaptive estimation of hysteresis thresholds. In *Proceedings of CVPR 1991*, pages 196–201, Hawaii, USA, June 1991.
- [9] J. Hayman. Multilevel thresholding for target tracking apparatus, 1985. US Patent, #4,550,435.
- [10] R. Kirby and A. Rosenfeld. A note on the use of (gray level, local average gray level) space as an aid in threshold selection. *IEEE Trans. Syst. Man Cybern.*, 9:860–864, 1979.
- [11] S. Z. Li. Markov random field modeling in computer vision, 1995. Springer-Verlag.
- [12] W. Lie. An efficient threshold-evaluation algorithm for image segmentation on spatial gray-level co-occurrences. *Signal Process.*, 33:121–126, 1993.
- [13] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equations of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1091, 1953.
- [14] V. Ramesh and R. Haralick. Automatic selection of tuning parameters for feature extraction sequences. In *Proceedings of CVPR 1994*, pages 672–677, Seattle, USA, June 1994.
- [15] M. Sezgin and B. Sankur. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, 13(1):147–168, January 2004.
- [16] S. German and D. German. Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Trans. on PAMI*, 6:721–741, 1984.
- [17] S. Venkatesh and P. Rosin. Dynamic threshold determination by local and global edge evaluation. *CVGIP: Graph. Models Image Process.*, 57:146–160, 1995.
- [18] A. Y. Wu, T. Hong, and A. Rosenfeld. Threshold selection using quadrees. *IEEE Trans. PAMI*, 4(1):90–04, 1982.
- [19] Y. J. Zhang. Two new concepts for N_{16} space. *IEEE Signal Processing Letters*, 6(9):221–224, 1999.